Planned Missingness with Multiple Imputation: an Application to Election Day Surveys.

René Bautista
NORC at the University of Chicago
University of Chicago
bautista-rene@norc.org

Marco A. Morales*
Wilf Family Department of Politics
New York University
marco.morales@nyu.edu

Abstract

This paper describes and links two conceptual ideas that have evolved in parallel in the fields of survey methodology and survey statistics. Such ideas have been empirically utilized by survey researchers with some disconnect until now. We argue that it is possible to align research strategies aimed to describe populations with strategies that seek to analyze them. Particularly, we unpack theoretical foundations that link the advantages of a survey methodology design for data collection, namely Planned Missingness (PM), with a single framework that takes advantage of a statistical approach known as Multiple Imputation (MI). When grounded on theory, models can be developed to go beyond data description and be used to test multivariate relationships. We discuss and illustrate this PM-MI pairing on an Election Day survey carried out in the 2006 Presidential election in Mexico. The exemplified advantages are applicable to broader settings where similar needs call for designs that cope with similar shortcomings.

Keywords: Planned Missingness, Multiple Imputation, Data Collection Methods, Exit Polls
Word count: 6,003

This paper extends a method introduced in Bautista, Morales, Abundis and Callegaro (2008). We thank Neal Beck, Alberto Díaz-Cayeros, Pat Egan, Federico Estévez, Andrew Gelman, Tobin Grant, Erik Hjermstad, Olivia Lau, Levente Littvay, Fernando Martel, Kristin Michelitch and Jonathan Nagler for useful comments to earlier versions of this paper. We owe a great debt of gratitude to Francisco Abundis at *Parametría* who supported the idea of testing PM-MI empirically in one of the most complicated - and close - elections in Mexican history.

*Authors are listed alphabetically as they share first authorship. Bautista conducted research for this project while affiliated to the Survey Research and Methodology (SRAM) at the University of Nebraska-Lincoln

In his now classic book "Survey Errors and Survey Costs", Groves (1989) distinguished two types of survey research users that are relevant to this paper: "describers" and "modelers". As per his portrayal, *describers* "use surveys to describe characteristics of a fixed population", while *modelers* "seek to identify causes of phenomena constantly occurring in a society". Over the couse of the previous decades, the literature has sought to improve the tools used by *describers*, namely tools for inferential statistics (*i.e.* the tasks of describers) such as sampling strategies and variance estimation (Valliant, Dever and Kreuter 2013), as well as questionnaire design and data collection methods (Tourangeau, Rips and Rasinski 2000; Dillman, Smyth and Christian 2009). For modelers, on the other hand, the literature has spent the last decades developing means to generate less biased estimates as a result of using more data that is better suited to nourish causal analysis (Lax and Phillips 2009; Ghitza and Gelman 2013; Sides and Vavreck 2013).

In an ideal world, both strains of the literature should have walked hand in hand, so that today researchers could have texts that were equally useful to describers and modelers. For example, it would be desirable to have a series of texts that combined data collection methods whose design were consistent with the assumptions made by robust regression models that would allow for a fuller exploitation of collected data.

Unfortunately, that is not the case, and both strains have developed somewhat independent of one another. To date, important work has been conducted in the field of survey methodology to improve data collection strategies aiming to reduce respondent burden (Mitofsky 2000; Fricker 2012) including an approach that can be described as *Planned Missingness* (PM). This approach consists of dividing the survey instrument into several sections. Each of those sections are administered to subset of respondents. Importantly, all respondents are administered a common set of questions. The outcome of such data collection strategy is a dataset with properties similar to a dataset with missing data; hence the name of planned missingness.

In parallel, deep progress has also been achieved in the field of survey statistics in developing techniques to conduct *Multiple Imputation* (MI), whereby missing data can be accounted for in the data analysis stage (Rubin 1987; Little and Rubin 2002). This technique can be loosely described as "simulating" missing data from the distribution that originates the full data distribution, by providing various "possible" values for each missing observation. These techniques have facilitated the analysis of datasets that otherwise would be difficult to conduct. But, to date, we have limited means to link advances in both bodies of literature (Peytchev 2012). A PM design can be undertaken, and also MI can be performed on a given dataset. Our contribution in this paper is to link the advantages of a survey methodology design - namely PM - into a single framework that takes advantage of a statistical approach - namely MI - to fully exploit the data collected on this design. We describe and illustrate this pairing with an example of an exit poll carried out in Mexico in the 2006 Presidential election. Nonetheless, the advantages that we exemplify here, are also applicable to broader settings where similar needs call for designs that cope with similar shortcomings.

1 Design Stage 1: Planning for Missingness

Planned Missingness is the name commonly given to survey designs in which the same target population is queried to answer different sets of questions, thus generating a controlled item non-response. For the design we propose here, different questionnaire versions are randomly administered to different subsamples of the population, but it is important to note that certain portions of the questionnaire are asked to all respondents, thus providing a baseline of information that is gathered from every unit in the general sample. As every individual in the sample answers a different questionnaire version, planned missingness generates various small-n data sets that contain certain pieces of information from a given population subset. Theoretically, as a result of random assignment, each data set should reflect the parameters of interest for the target population with a given level of uncertainty. Figure 1 provides a graphical description of the data collected using this planned missingness design.

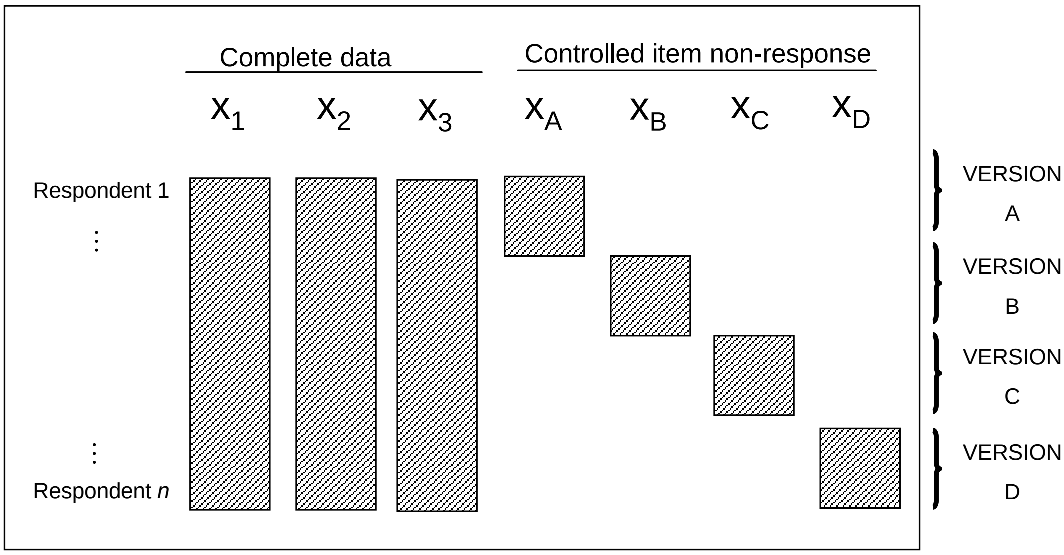

Figure 1: Missing data pattern generated with *Planned Missingness*

The mechanics of missingness that is planned. The key feature of this design is that data-missingness is not related to unobserved variables, a property commonly known in the statistical literature as *ignorability* (Rubin 1976, 1978). In our particular case, we go a step further with this design in making use of Rosenbaum and Rubin's (1983) notion of *strong ignorability*, whereby being "treated" with a questionnaire version ($I = 1$) depends on a variable created by the researcher ($I \in \{0, 1\}$) and, therefore, *known and fully observed.* As will become clear later, the consequences of this property will be crucial for the pairing we are proposing here. For now, let us simply note that the randomization in the survey design would permit the unbiased estimation of parameters of interest and that, for this feature to hold, it is necessary to assign the different questionnaire versions randomly to each population subset.

Stated more formally, since the assignment of a respondent to any questionnaire version (I) is random and no information in I changes our knowledge about the information collected by the survey ($\mathbf{x}$), we could state that

$$p(\mathbf{x}|I) = p(\mathbf{x}) \tag{1}$$

By extension, since any function of the observed information ($f(\mathbf{x})$) cannot help us predict the assignment to any questionnaire version either (as described in Dawid (1979), Theorem 2.1), we could extend eq. 1 to the following equivalence:

$$p(\mathbf{x}|I, f(\mathbf{x})) = p(\mathbf{x}|I) = p(\mathbf{x}) \tag{2}$$

In other words, *by construction* the information collected by a survey ($\mathbf{x}$) on a Planned Missingness design is *conditionally independent* ($\mathbf{x} \perp\!\!\!\perp I$) - in the Dawid (1979) sense - from the assignment mechanism (I), given any function of the data ($f(\mathbf{x})$). The usefulness of this statement will become evident shortly.

The empirics of missingness that is planned. Readers might justifiably ask themselves whether the estimated parameters from each subsample may not be themselves biased and impossible to validate since a particular questionnaire version is asked only to one sample subset. While we cannot *directly* verify the consistency of the estimates themselves, we can use an *indirect* approach to verify that the samples are similar amongst themselves with regards to the common information that they share.

It is helpful to think of each subset of the sample that is assigned a specific questionnaire version as "treated" with that particular questionnaire ($I = 1$). Evidently, a group "treated" with one questionnaire version is automatically "untreated" ($I = 0$) and hence a "control" for the rest of questionnaire versions. Consequently, given that the assignment of questionnaires is randomized, in expectation each of our subsamples would be balanced with regards to pre-treatment variables, which means that "within well-defined subgroups of treatment and control units, the distribution of covariates differ only randomly between the treatment and the control units" (Rubin 2008, pp. 809).

This feature - balance in pre-treatment covariates - can be validated empirically. Researchers can verify, for instance, by comparing sample moments, whether there is balance between subsamples. If that is not the case, empirical strategies could be applied. For example, Rosenbaum and Rubin (1983, 1984) define the *propensity score* as the conditional probability of assignment to treatment given covariates:

$$e(\mathbf{x}) = p(I = 1|\mathbf{x}) \tag{3}$$

They show that matching on it will balance the observed $\mathbf{x}$, but also that propensity scores are the coarsest among balancing scores (Rosenbaum and Rubin 1983, Theorem 2). Hence, were balance in pre-treatment covariates to fail, researchers could still use estimated propensity scores to eliminate systematic biases between the "treated" and "control" subsamples (Rosenbaum and Rubin 1985). This procedure may be thought of as analogous to post-stratification in traditional survey analysis (Kalton 1983), where adjustments to the data are performed with the use of weights to correct for biases created by data missingness.

Planned Missingness in other contexts. Planned missingness has been used for research purposes in various settings. For example Graham, Hofer and Piccinin (1994) used planned missingness as means to reduce omissions and attrition in the long questionnaires used by the Cancer Risk Behavior Survey, while still collecting sufficient information from subsamples of the surveyed population. Similarly, Littvay and Dawes (2007) apply it to ameliorate context effects in attitude questions, and randomly assign these questions to different respondents and estimate a latent variable that more accurately captures the construct under study. Also, Mitofsky (2000) used it in the context of exit polls to enhance the amount of information that can be gathered from voters in the brief interviews conducted as they leave polling stations. To the best of our knowledge, this is the general scope of what Planned Missingness has been used for in the literature so far.

2 Design Stage 2: Multiply Imputing the Missingness

Multiple Imputation is one among the available procedures devised to deal with missing data. Originally proposed by Rubin (1977, 1987), multiple imputation is a model-based approach to assign plausible values for missing data conditional on observed data. Briefly, the process consists of generating $m > 1$ data sets where no changes are made to the observed data, but $m > 1$ different plausible values are assigned to the missing data exploiting all observed information and the covariation among variables in the full data matrix. Figure 2 illustrates an intuitive way to think about this process.

For this design to work, ignorability in the missing data mechanism must be assumed (Rubin 1976). That is, missingness must not be conditional on unobservables and can be ignored given the appropriate conditioning on the observed data. It is important, at this point, to distinguish between data-missingness generated by the survey design, and item non-response that is independent from the survey design in this particular context. This distinction is discussed next.

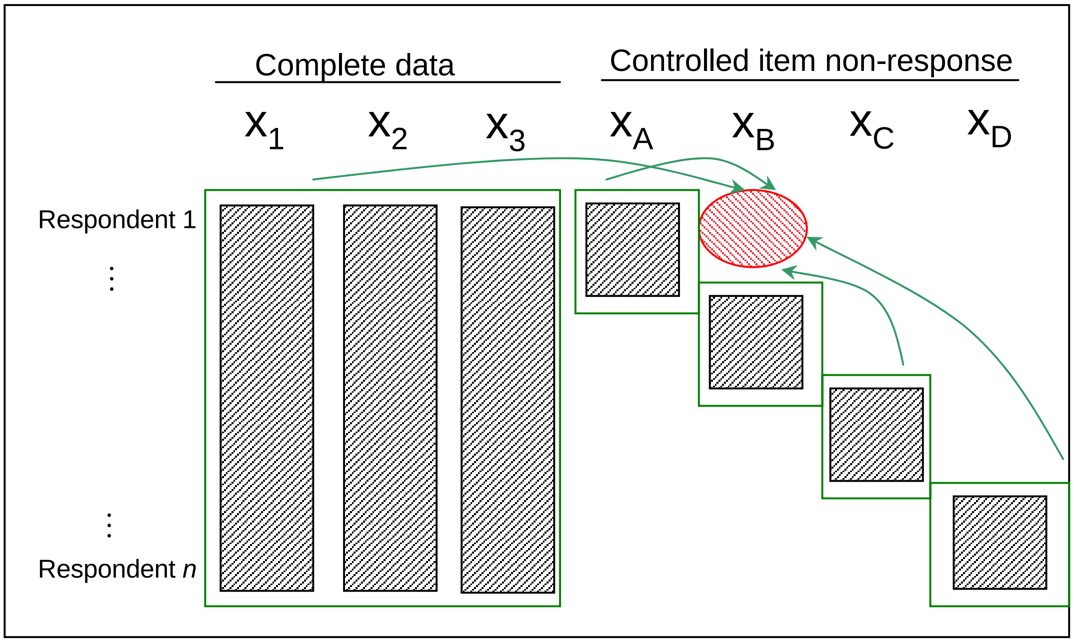

Figure 2: Missing data completion using *Multiple Imputation*

Planned Missingness vs item non-response. Planned Missingness, as is sometimes referred to in the literature, is related to a well-known concept in the survey methodology literature, namely item non-response. However, the main difference between them, is that Planned Missingness is due to the survey design - absent any choice from the respondent - while item non-response occurs when the respondent *chooses* not to answer the question. So, in essence, we end up with two conceptually different types of missing data; one *by design* and the other *by respondent's choice*.

Let us begin by characterizing a survey (S) by the two types of information inherent to it. On the one hand, we have information related to the survey design (D) that defines the way in which it is carried out. On the other hand, we have the responses (R) to every question included in the survey. Hence, each survey can be described as per

$$S \in \{D, R\} \tag{4}$$

In the case of survey design, we have information directly pertaining to the survey instrument as applied (D_{sur}), such as the questions that were included in the questionnaire, their phrasing, or their ordering, among others. Similarly, we have paradata for the survey (D_{para}), such as the race of the interviewer, the number of attempts made to contact the respondent, characteristics of non-respondents among others. Hence, if we group both elements together, we could describe the survey design by

$$D \in \{D_{sur}, D_{para}\} \tag{5}$$

In the case of survey responses, we must consider both observed (R_{obs}) or unobserved (R_{miss}) responses. In the first case (R_{obs}), we observe these data because respondents provided an answer to the questions they were asked. In the second case (R_{miss}), we fail to observe responses either because the questions were not asked or because respondents decided not to answer them. Hence survey responses could be described as

$$R \in \{R_{obs}, R_{miss}\} \tag{6}$$

With these simple definitions, we can embark in characterizations of data missingness (M), and distinguish the features that separate missingness due to survey design from item non-response, as well as those they have in common.

Characterizing Planned Missingness. In the case of PM, data-missingness results from questions not being asked to random subsets of the population. Therefore, the survey design (D_{sur}) is the only (observed) variable that can predict missingness. Formally,

$$P(M|D_{sur}) = P(M|R) \tag{7}$$

In other words, once we have conditioned on the survey design, there is no additional information on the responses - both observed and unobserved - that could help us characterize the mechanism that governs data missingness. That is, once we condition on the survey design (D_{sur}), Missingness is Completely At Random (MCAR). As defined by Rubin (1987), MCAR is a special case where observations are not only Missing at Random (MAR) but also Observed at Random (OAR). According to his definition, Missingness is MAR when the probability of missingness given observed and unobserved data, does not depend on *unobserved* data (Rubin 1987, Definition 1), and similarly it is CAR when for every value of the missing data, the probability of missingness given observed and unobserved data, does not depend on *observed* data (Rubin 1987, Definition 2). In the case of PM, this assumption is tenable because missingness depends only on whether respondents were randomly assigned to receive a specific question or not.

Characterizing item non-response. If we look at missingness from the respondent's end, we could conceive it as either *random* or *systematic*. While we recognize the effects on the efficiency of the estimation due to random item non-response, our focus here is on systematic non-response. Thus, we discuss systematic non-response in the sense that certain characteristics of the surveyed population can help predict data missingness in the survey. Think, for example, that education would be such characteristic, whereby less educated respondents fail to answer more complex questions. This would mean that if we condition on observed information on the survey (i.e. education), we could have missingness that is MAR, although in this case by assumption, and not by design. Formally,

$$P(M|R_{obs}) = P(M|R_{obs}, R_{miss}) = P(M|R) \tag{8}$$

In other words, once we condition on the observed information gathered by the survey, there would be no additional contribution from unobserved data to help us characterize the mechanism governing data missingness.

Finally, as a complement to these characterizations, we would have to describe Missingness Not at Random (MNAR) whereby no - observed or missing - data related to the survey (S) can assist to predict data missingness, thus making it unsuitable for MI. Formally,

$$P(M) = P(M|R, D) \tag{9}$$

That is, all information related to the survey cannot help us characterize the mechanism that governs missingness in the data.

Why Multiple Imputation works on Planned Missingness. For our purpose here, we are concerned with the data generating mechanism to the extent that it enables predicting missing data given observed data. That is, we care that ignorability can be assumed so that multiple imputation techniques are suitable (Rubin 1977, 1987). In the particular case proposed here, the observed data produced by Planned Missingness, and the estimated covariances for it, allow us to estimate plausible values for individuals who were not asked particular questions.

More specifically, the random assignment of the different questionnaire versions embedded in Planned Missingness guarantees ignorability, which is expected since missingness is the result of the survey design so that "almost all of the missingness is due to unasked questions" (Gelman, King and Liu 1998, pp. 847). We also assume that non-missing values in the data set are good predictors for the missing values, since they can be used to characterize distribution parameters for missing data. Furthermore, the design affords some additional desirable qualities: the sampling methodology is constant, and the questionnaire is applied on the same dates and by the same survey organization.

Multiply Imputing data generated with Planned Missingness. Certain features of Planned Missingness make the data particularly adequate for a proper imputation. Data-missingness, for example, is governed by the same mechanism across the data set: missingness is at random, and hence MAR. Also, item-nonresponse is governed by the same mechanism on each variable, which is MAR given covariates. Similarly, these features suggest that a covariance matrix common to all respondents can be reasonably assumed as they come from the same population surveyed at the same point in time. This is an issue that has worried researchers when dealing with imputation across surveys since incorrectly assuming a common covariance matrix could bias the imputations (Brehm 1998; Judkins 1998). Yet this should not be an issue in planned missingness for the reasons just mentioned.

Having determined that both types of missingness described on equations 7 and 8 fulfill the ignorability requisite by virtue of being MAR, usual analyses can be performed on each of these $m > 1$ completed data sets, and the results are combined using "Rubin Rules" (Rubin 1987). These rules produce asymptotically consistent and efficient estimates. This is because imputations carry a degree uncertainty with them that must also be incorporated in the estimates of the model. These rules suggests first to impute m values for each of the missing observations, producing m data sets where observed values do not change but missing ones take different plausible values that reflect the uncertainty on the imputation. Second, perform the usual analysis on each of the m data sets that, at this point, have imputed values where missing values existed. And third, use these m estimates to compute point estimates and variances for the parameters of interest. By virtue of this procedure, we are able to use all available data *as if* all questions had been asked to all respondents (Gelman et al. 1998) thus producing consistent and efficient estimates of the parameters of interest (Rubin 1987; King, Honaker, Joseph and Scheve 2001).

Alternatives to Multiple Imputation. Many possible methods have also been proposed to deal with missing data that could be applied to Planned Missingness situations: hot-deck imputation, cold-deck imputation, deductive imputation, conditional mean imputation, unconditional mean imputation, hot-deck random imputation, stochastic regression imputation, regression imputation, deductive imputation, exact match imputation, to name a few (Little 1992; Little and Rubin 2002; Weisberg 2005; Enders 2010).

The discussion above, does not imply that only MI is to be coupled with PM in this manner of research design. Theoretically, it would also be possible to use Maximum-Likelihood (ML) based methods to achieve similar results, a sort of PM-ML. Nonetheless, one recognized advantage of multiple imputation over other types of methods to deal with missing data - asides from its current popularity and ease of implementation - is the ability to reflect estimation uncertainty. This is a problem that needs to be addressed as single-value imputations would lead to underestimated variances (Rubin 1987). That is, instead of having one imputed value *as if it were the true value*, we can have $m > 1$ values from the predictive posterior distribution of the missing data. So, uncertain imputations will have high dispersions, while more certain ones will lie tightly around its expected value. With this information, variances are computed taking into account the within-estimates variance and the between-estimates variance producing efficient estimates with a limited number of imputations (Rubin 1987; King et al. 2001). In addition, MI is readily available in a variety of statistical softwares - in a way that ML is still not (Allison 2012) - and permits an easy inclusion of auxiliary information to model the missingness (Collins, Schafer and Kam 2001).

3 Motivating Example: PM-MI in an Election Day Survey

Among many other possible alternatives, a combination of Planned Missingness with Multiple Imputation can be applied on exit polls, with some important advantages for research purposes. Surveys carried out either before or after an election can suffer from important shortcomings; for example, voters who did not turn out on Election Day reporting that they did, or "true" voters misreporting who they actually favored. The latter is addressed directly by the exit poll design. People surveyed in exit polls are actual voters, which is a direct result of asking questions to individuals as they leave the polling places. It is cost-effective and conceptually more accurate to take a sample from a universe of actual voters (as is the case in exit polls) than it is to "filter" likely voters from the general public (as is the case in pre and post-election surveys). The former, which may be due to memory erosion, or social desirability mechanisms, can be minimized in exit polls. There is no incentive for respondents to appear as having voted for the winner of the election as they are still uncertain about the outcome of the election, especially in close elections. Also, as vote choice is asked minutes after the vote is cast, last-minute political events are unlikely to modify the self-reported vote choice, and they should remember their actual vote clearly.

Exit polls are a potentially exploitable instrument to analyze elections since they can overcome some limitations of pre- or post-election surveys when measuring attitudes and vote choice. The main challenge to enable exit polls to become a part of a researcher's tool kit is to design them so that they can collect substantial amounts of information without jeopardizing data quality. Lengthy questionnaires are typically associated to cognitive burden and lower response rates in exit polls (Mitofsky 1991; Moon 1999). However, exit polls are among the few opportunities to collect extensive data among actual voters. It should be evident by now that this condition may be achievable by combining the appropriate data collection design (*Planned Missingness*) with adequate statistical techniques (*Multiple Imputation*) as described in the previous sections.[1]

Before we proceed to our example, one important clarification is in place. Our aim in applying PM-MI to an exit poll is not to improve on the power of Election Day forecasts (*i.e.* predicting winners). Our aim is to improve on our ability to analyze and explain why votes were cast in the way they were by virtue of enabling regression estimators that are more efficient (in the statistical sense) and not biased by omitted variables. We proceed to illustrate the advantages of PM-MI in one of the most contested elections in recent times in Mexico: the 2006 Presidential election.

3.1 The Case of Mexico's Presidential Election Day Survey 2006

Motivating the use of PM-MI, contextual information. Immediately following the end of the campaign, five major hypotheses were advanced to explain the results of the 2006 Mexican Presidential election, whereby the candidate of the incumbent party - PAN's Felipe Calderón - won the election by a slim margin (Ugalde 2008). The first hypothesis states that the Federal Government aired a promotional campaign that focused on the achievements of the incumbent PAN President, Vicente Fox, referred to as the "continuity campaign", that was said to have boosted Felipe Calderón's candidacy. A second one poses that the PAN camp initiated an early negative campaign against the main challenger from the left - Andrés Manuel López Obrador - who would later retaliate; it was said that the negative campaign affected the PRD candidate. Third, it was believed that a federal program of conditional

[1] This notion on exit polls builds on Mitofsky's (2000) strategy to implement different questionnaires on exit polls as means to gather more information on general descriptions of opinion and attitudinal variables. Even though planned missingness has been implemented on exit polling designs, the data was never combined - nor multiply imputed - to obtain larger data sets to be used to analyze voting behavior more thoroughly.

cash-transfer social spending named "Oportunidades" favored the PAN candidate. Fourth, it was posited that the relatively good state of the economy, which implied no end-of-term economic crisis, had a positive impact on the PAN candidate. Finally, it was said that the high approval numbers of the outgoing President Fox produced coattails that helped the PAN candidate. With no comprehensive exit poll data it is hard to test such set of hypothesis *simultaneously*.

Implementing PM-MI. Planned missingness was implemented in an exit poll conducted by *Parametría*, one of the premier survey research firms in Mexico, for the 2006 Mexican Presidential election which collected information from 7,764 voters, with an approximate sampling error of +/- 1.1% with 95% of statistical confidence under simple random sampling.[2] This exit poll is the result of a stratified two-stage sampling design where 200 primary sampling units (*i.e.* precincts or "Electoral Sections") were selected as a nationwide sample with probability proportionate to size. The relative size of each cluster was the number of registered voters as determined by Mexico's Federal Electoral Institute (IFE). The number of surveys conducted on each one of these sampling units (between 7 and 70), depended on the precinct size and turnout rate. Since exit poll designs typically do not generate data used for traditional calculations of response rates, the closest estimate in line with AAPOR's standards can be generated by a variant of Slater and Christensen's (2002) $RR5_s$, which renders a response rate of 53%.[3]

In particular, a mixed mode data collection method was implemented under a missing-by-design election day survey. In the context of exit polls, a mixed mode data collection method is the most suitable option for populations with low literacy level, such as Mexico

[2] The square root of the design effect (DEFF) for each of the vote choice variable was on average 1.98, thus the confidence intervals are approximately 2 times as large as they are under simple random sampling.

[3] Slater and Christensen's (2002) rate is defined as $RR5_s = I/[(I+P)+R]$ where I is the number of completed interviews, P is the number of partial interviews, and R is the number of refusals. In our case, $P = 0$ as all questionnaires are considered as completed.

(Bautista, Callegaro, Vera and Abundis 2007). Specifically, the interviewer first approaches selected respondents and collects their demographic information and presidential approval. Immediately following, the interviewer hands a blank facsimile of the official ballot to the respondent who deposits it in a portable ballot box.[4] Finally, four different versions of the last portion of the questionnaire were administered rotatively with each version differing on the additional information that was collected.[5] In particular, Version "A" collected respondents' recollection of having watched President Vicente Fox administration's campaign advertisements, and whether they are beneficiaries of several social policy programs.[6] Version "B" asked respondents to evaluate whether each candidate and party ran a negative or positive campaign.[7] Version "C" asked respondents to place candidates, parties and themselves on a 7-point ideological scale, along with their party identification and evaluations of the state of the economy.[8] Version "D" did not gather any additional information relevant to this analysis, but is useful for a more precise estimation of the covariance matrix of the data, and hence for a more precise imputation. Given the high missingness in the Planned Missingness-Multiple Imputation data, $m = 10$ data sets were imputed that "filled-in" the missing values on 37 variables. (See online Appendix for details.)

Applying theoretical tools to analyze data generated with PM-MI. As a result of applying the PM-MI design to *Parametría*'s exit poll it was possible to collect sufficient information to evaluate the plausibility of the main hypotheses advanced to explain the results of the 2006 election. A multinomial probit analysis was performed on the data to test them. For ease of exposition, we present estimates graphically as suggested by Gelman, Pasarica and Dodhia (2002). Figure 3 shows simulations of changes in probabilities - also

[4] A control number is printed on each ballot facsimile that allows matching reported vote choice with the information collected in the questionnaires.

[5] Interviewers were given the four versions of the questionnaire in a presorted order.

[6] 2,032 voters received this version leaving 5,732 answers to be imputed

[7] 1,859 voters answered this version leaving 5,905 answers to be imputed.

[8] 1,795 voters replied to this version, leaving 5,969 answers to be imputed.

known as first differences (King 1998) - of a typical individual voting for candidate j given variations on a particular variable (see online Appendix for full details on the estimation, and simulations).

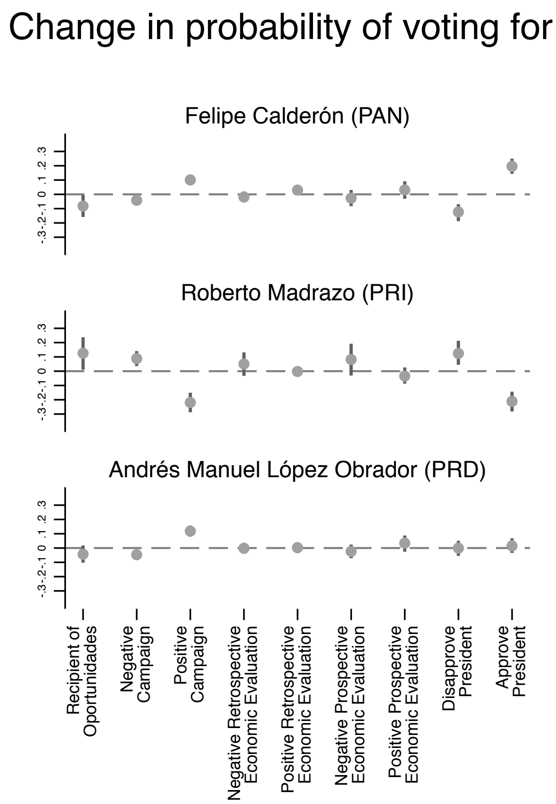

Figure 3: First differences - and their associated 95% confidence intervals - on the simulated probability of voting a candidate given a change in the specified variable. Simulations are generated for the probability of voting for the candidate denoted on each graph comparing a "typical" individual with the lowest value for the specified variable to a "typical" individual that has the highest value on that same variable, holding all other variables constant.

Briefly, the empirical analysis shows that, net of other factors, only one of the hypothesis finds empirical support; namely, the single most important predictor for voting for Calderón was approval of President Fox. To date, scholarly research on the 2006 Mexican election has used pre- and post-election, as well as exit poll survey data when seeking to explain the outcome of the election (Moreno 2007; Estrada and Poiré 2007; Moreno and

Méndez 2007; Moreno 2009; Beltrán 2009*a*,*b*; Lehoucq 2009; Guardado 2009; Singer 2009; Abundis and Ley 2009). However, neither of these analyses has explored *simultaneously* all the factors that are thought to have determined the outcome of the 2006 Presidential election in Mexico. This is mostly due to the limited data available from the surveys used in these studies. In this context, our proposed approach is able to shed light on the determinants of vote choice in the 2006 election by virtue of the use of PMMI.[9]

4 Discussion

Readers versed in experimental traditions might recognize the effects of a Planned Missingness design as analogous to the "Fundamental Problem of Causal Inference" (Holland 1986), namely that we cannot simultaneously observe the responses of treated (when asked a question subset) and the untreated (when not asked a question subset) respondents, thus preventing us from making inferences at the respondent level. Nevertheless, by virtue of the combination of a *Planned Missingness* design with *Multiple Imputation* techniques, we can draw power from the fact that each one of the respondents on each sample initially had an equal probability of being selected into the sample, and that probability was then translated into an intent to assign them to one version of the questionnaire. By design, we can only observe their answers to that specific questionnaire, and have the answers to all other versions as missing values. But even if we do not know how the answers that each person would have given to other versions of the questionnaire, based on the rest of the respondents who actually answered those versions we are able to estimate population parameters from the available information. In addition, this property makes this design particularly well suited for designing experiments with an aim to generate causal inference.

[9]Results deserve further discussion, but that falls out of the scope of this paper, which is to present and justify the use of PMMI, illustrating the case with an exit poll.

We recognize that an imputation is only as good as the correlation between the observed and the missing covariates: the better the correlation across these variables, the more accurate and efficient the imputation will be (Brehm 1998; Binder 1998). Further research is needed to improve the design in order to enhance correlations across variables.

Optimal patterns of planned missingness. One way to improve the quality of the imputation is through the patterns of the planned missingness. In our implementation of PM-MI in the 2006 Mexican election, we chose to create planned missingness with question blocks. That is, *questions that were not asked to all respondents* were only included in one questionnaire version (a block), with no overlaps across questionnaires. Other split-block designs where questions overlap in "Swiss-cheese" missing-data patterns (Judkins 1999) are also possible. Graham, Hofer and MacKinnon (1996) show that estimates that use data from unique block designs are as efficient as those generated with split-block designs, although efficiency might be better in split-block designs depending on the correlations between and across questions in a block. They also show that estimates using data from question block designs become more efficient as the correlation of the questions within the block increases. Similarly, the efficiency of the estimates based on split-block design data depend highly on the correlations between blocks of questions, a rationale that is supported by Raghunathan and Grizzle (1995). Therefore, it seems to be good practice to group blocks of questions in a way such that correlation is enhanced: between the questions in a block if grouping questions in unique blocks, or across blocks of questions if using split-block designs.

In view of the potential limitations of our design, it is useful to recount our reasons for choosing it. The first, and perhaps most obvious reason, is that grouping blocks of questions by version is *logistically* much simpler to implement in the field. From a practitioner standpoint, it is paramount to avoid adding sources of confusion to the data collection protocol. The second reason is to meet various clients' needs using a syndicated survey, which considerably reduces its cost. Instead of fielding different exit polls, different versions of a questionnaire were fielded out keeping a set of variables common to all questionnaires for consistency-verification purposes.[10] Note also that Graham, Taylor, Olchowski and Cusmille (2006) favor this strategy on logistical grounds alone.

Optimal number of questions and questionnaire versions, future research. An additional aspect that should be further investigated is the number of questionnaires that would be optimal in the planned missingness design in order to have "good" (*i.e.* efficient and consistent) imputations. Also, further research is needed on the number of questions to include on each questionnaire version.[11] There are two possible ways to answer these questions. First, holding the sample size constant, the number of questions and/or questionnaires is related to the algorithm employed to impute. In other words, we need to find out what is the lower bound for the number of variables to be used in the imputation that still produce efficient imputations. Simulations might provide useful guidance on this matter. Second, the number of questions and/or questionnaires is closely related to sample size in the exit poll. The larger the sample size, the more questions and/or questionnaires could be included. Yet there is no standard "optimal" sample size for exit polls, as it is typically determined on a case-by-case basis as long as the final number of sampling observation units (*i.e.* voters) may vary as a function of turnout and response rates.[12]

[10] All things considered, it might have been a better alternative to use a split-block design, although this is a more challenging alternative to implement for logistic reasons. That said, the distribution of questions within each questionnaire version does enhance the correlation across variables, as questions are grouped by topic.

[11] Not all questionnaires must have the same sample size, and it might make sense to have a particular block with a larger relative sample size if the question under investigation justifies this choice.

[12] Performing full information maximum likelihood (FIML) estimation, Littvay (2009) finds that an increase in the number of questionnaire versions does not increase bias in the estimators, although it may decrease their statistical power. Hence he suggests increasing sample size to offset the loss in statistical power.

5 Conclusions

In an ideal world with unlimited resources and very patient respondents, we would have exhaustive instruments where respondents were administered comprehensive questionnaires. But we cannot always do what is ideal, so we need to do what makes practical sense. This paper is an attempt at doing that, by linking progress in the fields of survey methodology and survey statistics empirically, while also drawing the theoretical parallels on both approaches.

To date, the literature has been limited in aligning research strategies to describe populations with those that seek to analyze them. Furthermore, the limitations in the existing literature extend to generating predictive models that stem from these exercises. Throughout this paper, we have argued that it is possible to do so when, grounded on theory, we develop models that aims to go beyond data description, and into more analytical/predictive grounds. We have also unpacked the theoretical foundations that link them both in practice, and that show under what conditions they may be best suited to fulfill broader research objectives.

We have exemplified this approach with an exit polling design that combines Planned Missingness and Multiple Imputation; concepts that have been advanced previously in the literature but, to the best of our knowledge, never fully implemented in an exit polling context. As the illustration from the Mexican 2006 election shows, the design does not seem to generate particular problems with the quality of the data being collected. More importantly, it enables a much richer data analysis, thus producing more reliable and consistent estimates for voting behavior analysis.

References

Abundis, F., and Ley, S. (2009), "Votos y votantes en la elección federal de 2006," *Política y Gobierno*, Temático, 177–199.

Allison, P. D. (2012), Handling Missing Data by Maximum Likelihood,. Keynote address at SAS Global Forum.

Bautista, R., Callegaro, M., Vera, J. A., and Abundis, F. (2007), "Studying Nonresponse in Mexican Exit Polls," *International Journal of Public Opinion Research*, 19(4), 492–503.

Bautista, R., Morales, M. A., Abundis, F., and Callegaro, M. (2008), "Exit polls as valuable tools to understand voting behavior: Using an advanced design in Mexico - Measurement Error," in *Elections and Exit polling*, eds. F. Scheuren, and W. Alvey, New York, NY: John Wiley & Sons.

Beltrán, U. (2009*a*), "Ideología y polarización en la elección de 2006," *Política y Gobierno*, Temático, 82–106.

Beltrán, U. (2009*b*), "Publicidad y preferencias," *Política y Gobierno*, Temático, 236–271.

Binder, D. A. (1998), "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys: Comment," *Journal of the American Statistical Association*, 93(443), 858–859.

Brehm, J. (1998), "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys: Comment," *Journal of the American Statistical Association*, 93(443), 859–860.

Collins, L. M., Schafer, J. L., and Kam, C.-M. (2001), "A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures," *Psychological Methods*, 6(4), 330–351.

Dawid, A. P. (1979), “Conditional Independence in Statistical Theory,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1), 1–31.

Dillman, D. A., Smyth, J. D., and Christian, L. M., eds (2009), *Internet, mail, and mixed-mode surveys: the tailored design method*, 3rd edn, Hoboken, NJ: Wiley & Sons.

Enders, C. K. (2010), *Applied Missing Data Analysis*, New York, NY: Guilford Press.

Estrada, L., and Poiré, A. (2007), “Taught to protest, learning to lose,” *Journal of Democracy*, 18(1), 73–87.

Fricker, S. (2012), Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse and Respondent Burden,. Paper presented at the 67th Annual Meeting of the American Association for Public Opinion Research.

Gelman, A., King, G., and Liu, C. (1998), “Not Asked and Not Answered: Multiple Imputation for Multiple Surveys,” *Journal of the American Statistical Association*, 93(443), 846–857.

Gelman, A., Pasarica, C., and Dodhia, R. (2002), “Let’s Practice What We Preach: Turning Tables into Graphs,” *The American Statistician*, 56(2), 121–130.

Ghitza, Y., and Gelman, A. (2013), “Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Groups,” *American Journal of Political Science*, 57(3), 762–776.

Graham, J. W., Hofer, S. M., and MacKinnon, D. P. (1996), “Maximizing the Usefulness of Data Obtained with Planned Missing Value Patterns: An Application of Maximum Likelihood Procedures,” *Multivariate Behavioral Research*, 31(2), 197–218.

Graham, J. W., Hofer, S. M., and Piccinin, A. M. (1994), "Analysis with missing data in drug prevention research," in *Advances in data analysis for prevention intervention research*, eds. L. M. Collins, and L. Seitz, Washington, DC: National Institute on Drug Abuse.

Graham, J. W., Taylor, B. J., Olchowski, A. E., and Cusmille, P. E. (2006), "Planned Missing Data Designs in Psychological Research," *Psychological Methods*, 11(4), 323–343.

Groves, R. (1989), *Survey errors and survey costs*, New York, NY: John Wiley & Sons.

Guardado, J. (2009), "La identidad partidista en México: Las dimensiones políticas de la competencia en las elecciones presidenciales de 2000 y 2006," *Política y Gobierno*, Temático, 137–175.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396), 449–484.

Judkins, D. R. (1998), "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys: Comment," *Journal of the American Statistical Association*, 93(443), 861–864.

Judkins, D. R. (1999), Imputing for Swiss cheese patterns of missing data,. Proceedings of Statistics Canada Symposium 97, New Directions in Surveys and Censuses.

Kalton, G. (1983), *Compensating for Missing Survey Data (Research Report Series / Institute for Social Research)*, Ann Arbor, MI: University of Michigan Press.

King, G. (1998), *Unifying Political Methodology. The Likelihood Theory of Statistical Inference*, Ann Arbor, MI: University of Michigan Press.

King, G., Honaker, J., Joseph, A., and Scheve, K. (2001), "Analyzing Incomplete Political Science Data: An alternative algorithm for multiple imputation," *American Political Science Review*, 95(1), 49–69.

Lax, J. R., and Phillips, J. H. (2009), "How Should We Estimate Public Opinion in The States," *American Journal of Political Science*, 53(1), 107–121.

Lehoucq, F. (2009), "Ingreso y cálculos electorales en 2006," *Política y Gobierno*, Temático, 107–136.

Little, R. J. (1992), "Regression with Missing X's: a Review," *Journal of the American Statistical Association*, 87(420), 1227–1237.

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, 2nd edn, New York, NY: John Wiley & Sons.

Littvay, L. (2009), "Questionnaire Design and Considerations with Planned Missing Data," *Review of Psychology*, 16(2), 103–113.

Littvay, L., and Dawes, C. T. (2007), Alleviation of Context Effects in Attitude Questions Through Computer Assisted Interviewing and a Planned Missingness Data Design,. Presented at the 2nd European Survey Research Association Conference, Prague, Czech Republic.

Mitofsky, W. J. (1991), "A short history of exit polls.," in *Polling and presidential election coverage*, eds. P. J. Lavrakas, and J. K. Holley, Newbury Park, CA: Sage.

Mitofsky, W. J. (2000), "At the polls: Mexican democracy turns the corner," *Public Perspective*, 11(5), 37–40.

Moon, N. (1999), "Exit polling: a special case," in *Opinion polls: History, theory and pratice*, Manchester: Manchester University Press.

Moreno, A. (2007), "The 2006 Mexican Presidential Election: The Economy, Oil Revenues, and Ideology," *PS: Political Science & Politics*, 40(1), 15–19.

Moreno, A. (2009), *La Decisión Electoral. Votantes, Partidos y Democracia en México.*, Mexico, DF: Miguel Ángel Porrúa.

Moreno, A., and Méndez, P. (2007), "La identificación partidista en las elecciones presidenciales de 2000 y 2006 en México: ¿Desalineación o realineación?," *Política y Gobierno*, XIV(1), 43–75.

Peytchev, A. (2012), "Multiple Imputation for Unit Nonresponse and Measurement Error," *Public Opinion Quarterly*, 76(2), 214–237.

Raghunathan, T. E., and Grizzle, J. E. (1995), "A Split Questionnaire Survey Design," *Journal of the American Statistical Association*, 90(429), 54–63.

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of Propensity Score in Observational Studies for Causal Inference," *Biometrika*, 70(1), 41–55.

Rosenbaum, P. R., and Rubin, D. B. (1984), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79(387), 516–524.

Rosenbaum, P. R., and Rubin, D. B. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistician*, 39(1), 33–38.

Rubin, D. B. (1976), "Inference and Missing data," *Biometrika*, 63(3), 581–592.

Rubin, D. B. (1977), "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, 72(359), 538–543.

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6(1), 34–58.

Rubin, D. B. (1987), *Multiple imputation for nonresponse in surveys*, New York, NY: Wiley & Sons.

Rubin, D. B. (2008), “For Objective Causal Inference, Design Trumps Analysis,” *Annals of Applied Statistics*, 2(3), 808–840.

Sides, J., and Vavreck, L. (2013), *The Gamble:Choice and Chance in the 2012 Presidential Election*, Princeton, NJ: Princeton University Press.

Singer, M. M. (2009), “”Defendamos lo que hemos logrado”: El voto económico en México durante la elección presidencial de 2006,” *Política y Gobierno*, Temático, 199–235.

Slater, M., and Christensen, H. (2002), Applying AAPOR’s Final Disposition Codes and Outcome Rates to the 2000 Utah College’s Exit Poll,. Presented at the 57th Annual Meeting of the American Association for Public Opinion Research, St. Pete Beach, Florida.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000), *The Psychology of Survey Response*, New York, NY: Cambridge University Press.

Ugalde, L. C. (2008), *Así lo viví*, Mexico, DF: Grijalbo Mondadori.

Valliant, R., Dever, J. A., and Kreuter, F. (2013), *Practical Tools for Designing and Weighting Survey Samples*, New York, NY: Springer.

Weisberg, H. F. (2005), *The Total Survey Error Approach. A Guide to the New Science of Survey Reserach*, Chicago, IL: University of Chicago Press.